

IFW

PTO/SB/21 (09-04)

**TRANSMITTAL  
FORM**

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

1

Application Number

10/754,164

Filing Date

January 9, 2004

First Named Inventor

NAKAGAWA, Yoshihito

Art Unit

2171

Examiner Name

Unassigned

Attorney Docket Number

16869P-102400US

**ENCLOSURES (Check all that apply)**

<input type="checkbox"/> Fee Transmittal Form <input type="checkbox"/> Fee Attached <input type="checkbox"/> Amendment/Reply <input type="checkbox"/> After Final <input type="checkbox"/> Affidavits/declaration(s) <input type="checkbox"/> Extension of Time Request <input type="checkbox"/> Express Abandonment Request <input type="checkbox"/> Information Disclosure Statement  <input checked="" type="checkbox"/> Certified Copy of Priority Document(s)  <input type="checkbox"/> Reply to Missing Parts/ Incomplete Application <input type="checkbox"/> Reply to Missing Parts under 37 CFR 1.52 or 1.53	<input type="checkbox"/> Drawing(s) <input type="checkbox"/> Licensing-related Papers <input type="checkbox"/> Petition <input type="checkbox"/> Petition to Convert to a Provisional Application <input type="checkbox"/> Power of Attorney, Revocation Change of Correspondence Address <input type="checkbox"/> Terminal Disclaimer <input type="checkbox"/> Request for Refund <input type="checkbox"/> CD, Number of CD(s) _____ <input type="checkbox"/> Landscape Table on CD	<input type="checkbox"/> After Allowance Communication to TC <input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences <input type="checkbox"/> Appeal Communication to TC (Appeal Notice, Brief, Reply Brief) <input type="checkbox"/> Proprietary Information <input type="checkbox"/> Status Letter <input checked="" type="checkbox"/> Other Enclosure(s) (please identify below): Return Postcard
<b>Remarks</b> The Commissioner is authorized to charge any additional fees to Deposit Account 20-1430.		

**SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT**

Firm Name	Townsend and Townsend and Crew LLP		
Signature			
Printed name	Chun-Pok Leung		
Date	December 10, 2004	Reg. No.	41,405

**CERTIFICATE OF TRANSMISSION/MAILING**

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature			
Typed or printed name	Joy Salyador	Date	December 10, 2004

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日            2 0 0 3 年   6 月   3 日  
Date of Application:

出 願 番 号            特 願 2 0 0 3 - 1 5 8 2 7 1  
Application Number:  
[ST. 10/C] :            [ J P 2 0 0 3 - 1 5 8 2 7 1 ]

出      願      人            株式会社日立製作所  
Applicant(s):

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2 0 0 4 年   2 月 1 9 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫

出証番号    出証特 2 0 0 4 - 3 0 1 1 0 5 8

【書類名】 特許願

【整理番号】 HI030203

【提出日】 平成15年 6月 3日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 中川 義仁

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 小笠原 裕

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 内海 勝広

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 中山 信一

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 ▲高▼田 豊

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

## 【代理人】

【識別番号】 110000176  
【氏名又は名称】 一色国際特許業務法人  
【代表者】 一色 健輔

## 【手数料の表示】

【予納台帳番号】 211868  
【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ストレージ制御装置の制御方法及びストレージ制御装置

【特許請求の範囲】

【請求項 1】 情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、

第 1 の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第 1 の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第 2 の前記チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項 2】 情報処理装置から送信されるデータ入出力要求を受信して前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、前記各チャンネル制御部からアクセス可能な共有メモリを備えるストレージ制御装置の制御方法であって、

第 1 の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第 1 の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合に当該情報処理装置を特定可能な識別子を前記共有メモリに記憶し、

第 2 の前記チャンネル制御部が、適宜なタイミングで前記共有メモリにアクセスすることにより前記識別子を取得し、

前記第 2 のチャンネル制御部が、取得した前記識別子を有する前記情報処理装置

から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項 3】 情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、前記各チャンネル制御部からアクセス可能な共有メモリを備えるストレージ制御装置の制御方法であって、

第 1 の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第 1 の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合に当該情報処理装置を特定可能な識別子を前記共有メモリに記憶するとともに前記障害が生じている旨を第 2 の前記チャンネル制御部に通知し、

前記第 2 のチャンネル制御部が、前記通知を受信したのに応じて前記共有メモリにアクセスすることにより前記識別子を取得し、

前記第 2 のチャンネル制御部が、取得した前記識別子を有する前記情報処理装置から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項 4】 請求項 1 に記載のストレージ制御装置の制御方法において、前記データ入出力要求に応じて実行される処理を制限するための処理は、障害が検知された前記情報処理装置から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて出力される前記制御信号を出力しないように制御する処理であること、を特徴とするストレージ制御装置の制御方法。

【請求項 5】 請求項 1 に記載のストレージ制御装置の制御方法において、前記第 1 のチャンネル制御部は、前記チャンネル制御部と前記情報処理装置との間でハートビート信号を通信し、前記ハートビート信号が途絶したことをもって前記情報処理装置に障害が生じていることを検知すること、を特徴とするストレージ

ジ制御装置の制御方法。

【請求項 6】 請求項 1 乃至 5 のいずれかに記載のストレージ制御装置の制御方法において、

前記第 1 のチャンネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてファイル指定によるデータ入出力要求を受け付ける機能を備え、前記第 2 のチャンネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてブロック指定によるデータ入出力要求を受け付ける機能を備えること、を特徴とするストレージ制御装置の制御方法。

【請求項 7】 情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、

前記情報処理装置から送信されてくる要求に応じて前記情報処理装置にファイル管理情報を送信し、

前記情報処理装置から送信されてくる前記ファイル管理情報に基づいて生成されたデータ入出力要求を受信してこれに応じたデータの書き込み／読み出しを前記記憶デバイスに対して実行し、

第 1 の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

前記第 1 のチャンネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第 2 の前記チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項 8】 請求項 1 に記載のストレージ制御装置の制御方法において、前記情報処理装置と通信することにより行われる前記情報処理装置の動作の監視は、前記情報処理装置と前記第 1 のチャンネル制御部において動作しているクラ

ソフトウェアの機能により実現されること、  
を特徴とするストレージ制御装置の制御方法。

【請求項 9】 請求項 1 に記載のストレージ制御装置の制御方法において、  
前記第 1 のチャンネル制御部は LAN に接続するためのインタフェースを有し、  
前記第 2 のチャンネル制御部は SAN に接続するためのインタフェースを有する  
こと、を特徴とするストレージ制御装置の制御方法。

【請求項 10】 情報処理装置から送信されるデータ入出力要求を受信して、  
前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み  
出しを制御するための制御信号を出力する、互いに通信可能に接続される複数の  
チャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込  
み／読み出しを行うディスク制御部と、

第 1 の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情  
報処理装置の動作を監視する手段と、

第 1 の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合  
に、前記情報処理装置から送信されたデータ入出力要求が着信される第 2 の前記  
チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限する  
ための処理を実行する手段と、

を備えることを特徴とするストレージ制御装置。

【請求項 11】 請求項 10 に記載のストレージ制御装置において、  
前記第 1 のチャンネル制御部は、前記情報処理装置から送信される前記データ入  
出力要求としてファイル指定によるデータ入出力要求を受け付ける機能を備え、  
前記第 2 のチャンネル制御部は、前記情報処理装置から送信される前記データ入出  
力要求としてブロック指定によるデータ入出力要求を受け付ける機能を備えるこ  
と、を特徴とするストレージ制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、ストレージ制御装置の制御方法及びストレージ制御装置に関する

。



**【 0 0 0 2 】****【従来の技術】**

I T 関連産業の発達に伴い、コンピュータシステムが取り扱うデータ量が飛躍的に増加している。かかる膨大なデータを効率よく利用し管理するために、データセンタ等においては、ディスクアレイ装置等などの記憶装置と情報処理装置とを専用のネットワークで接続し、ディスクドライブ等のストレージに対する高速かつ大量なアクセスを実現するようにしたストレージシステムが構築されている。上記ネットワークとしては高速なデータ転送を実現するために、ファイバチャネルプロトコルに従った通信機器を用いて構築されるネットワークである、S A N (Storage Area Network) が用いられるのが一般的である。また、最近では、複数の記憶装置と情報処理装置とを接続する上記ネットワークとして T C P / I P (Transmission Control Protocol/Internet Protocol) プロトコルを用いた L A N (Local Area Network) を用い、情報処理装置からファイル指定によるアクセスを実現した N A S (Network Attached Storage) と呼ばれる装置も開発されている。

**【 0 0 0 3 】**

昨今では、ストレージシステムが取扱うデータに対する社会的な重要性が増しており、ストレージシステムには高い信頼性や可用性が求められるようになってきている。ここでストレージシステムの信頼性や可用性を向上させるための技術として、クラスタと呼ばれる技術が知られている。例えば、複数の情報処理装置でクラスタを構成し、情報処理装置間で障害の監視と障害検出時のフェールオーバーを実現することで、ストレージシステムの信頼性や可用性を向上させることができる。

**【 0 0 0 4 】****【特許文献 1】**

特開 2 0 0 2 - 3 5 1 7 0 3 号公報

**【 0 0 0 5 】****【発明が解決しようとする課題】**

ところで、ストレージシステムにおいては、例えば、情報処理装置側で障害が

発生しているにもかかわらず記憶装置側でそのまま情報処理装置から送信されるデータ入出力要求についての処理を続けていると、記憶装置に記憶されているデータの破損やデータに管理上の不整合等の問題が生じる可能性がある。従って、ストレージシステムの信頼性や可用性をより確実に確保するためには情報処理装置側で発生した障害の影響ができるだけ記憶装置側に及ばないようにすることも重要である。

#### 【0006】

本発明は上記課題を鑑みてなされたものであり、ストレージシステムの信頼性や可用性を向上させることができる、ストレージ制御装置の制御方法及びストレージ制御装置を提供することを主たる目的とする。

#### 【0007】

##### 【課題を解決するための手段】

上記目的を達成する本発明の主たる発明は、情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、第1の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、第1の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第2の前記チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行することとする。

#### 【0008】

前記制御信号は、例えば、直接的にもしくは後述する共有メモリやキャッシュメモリを介して間接的に、前記チャンネル制御部から前記ディスク制御部に対して伝達される信号である。情報処理装置において障害が発生すると、そのことが第1のチャンネル制御部により検知され、これに応じて第2の前記チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理が実行

される。これにより障害が生じている情報処理装置から送られてくる異常なデータ入出力要求に応じて第2のチャネル制御部が異常な処理を行うことにより記憶デバイスに記憶されているデータを破損したり、データに管理上の不整合が生じたりすることが防止され、ストレージシステムの信頼性や可用性を向上させることができる。

#### 【0009】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄及び図面により明らかにされる。

#### 【0010】

##### 【発明の実施の形態】

図1に本発明の一実施例として説明するストレージシステムの概略構成を示している。このストレージシステムは、例えば、データセンタ等において構築される。ストレージシステムは、例えば、インターネット上のWebサイトの運用、ASPサービスの提供、銀行や証券会社等におけるオンラインシステムやバッチ処理システムの運用等、各種の用途に利用される。

#### 【0011】

ストレージシステムは、ストレージ制御装置100、ストレージ制御装置100に通信可能に接続しストレージ制御装置100に対してデータの書き込み／読み出しに関する要求（以下、データ入出力要求と称する）を送信する情報処理装置1乃至2（200）、ストレージ制御装置100により制御されデータの記憶領域を提供する1台以上のディスクドライブ300（記憶デバイス）を含んで構成される。ストレージ制御装置100は、情報処理装置200から送信されてくるデータ入出力要求を受信して、前記データ入出力要求に応じてディスクドライブ300に記憶されているデータの書き込みや読み出しに関する処理を行う。

#### 【0012】

ストレージ制御装置100と情報処理装置1乃至2（200）とは、LAN（Local Area Network）400を介して通信可能に接続されている。LAN400を介して行われる通信は、例えば、TCP/IPプロトコルに従って行われる。情報処理装置1乃至2（200）とストレージ制御装置100とは、SAN（St

orage Area Network) 500 を介して通信可能に接続されている。SAN500 を介して行われる通信は、例えば、ファイバチャネル、iSCSI、FICON (登録商標)、ESCON (登録商標)、ACONARC (登録商標)、FIBARC (登録商標) 等に従って行われる。

#### 【0013】

情報処理装置 1 乃至 2 (200) は、いずれも CPU (Central Processing Unit) やメモリを備えるコンピュータである。情報処理装置 1 乃至 2 (200) は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等である。情報処理装置 1 乃至 2 (200) が備える CPU により各種プログラムが実行されることにより、情報処理装置 1 乃至 2 (200) が提供する様々な機能が実現される。情報処理装置 200 では、オペレーティングシステムが動作している。オペレーティングシステム上では、各種のアプリケーションソフトウェアが動作している。後述するクラスタソフトウェアも、上記オペレーティングシステム上で動作するアプリケーションソフトウェアの一つである。

#### 【0014】

図 2 に情報処理装置 200 の典型的なハードウェア構成を示している。情報処理装置 200 は、CPU 211、メモリ 212、記録媒体読取装置 214、入力装置 215、出力装置 216、記憶装置 217、LAN インタフェース 218、SAN インタフェース 219、等を備えて構成される。CPU 211 は、メモリ 212 に格納されているプログラムを実行する。記録媒体読取装置 214 は、記録媒体 230 に記録されているプログラムやデータを読み取るための装置である。記録媒体 230 としては、例えば、フレキシブルディスクや CD-ROM、DVD-ROM、DVD-R、DVD-RW、半導体メモリ等が用いられる。記録媒体読取装置 214 は、情報処理装置 200 に内蔵される形態とすることもできるし、外付されている形態とすることもできる。

#### 【0015】

記憶装置 217 は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置、等である。入力装置 215 は、ユーザやオペレータ等の人による情報処理装置 200 へのデータ入力等のためのユーザインタフェースである。

入力装置 215 としては、例えば、キーボードやマウス等が用いられる。出力装置 216 は、例えば、ディスプレイやプリンタ等である。

#### 【0016】

LAN インタフェース 218 は、情報処理装置 200 を LAN 400 に接続するための通信インタフェースである。LAN インタフェース 218 としては、例えば、イーサネット（登録商標）に対応したネットワークカードが用いられる。SAN インタフェース 219 は、情報処理装置 200 を SAN 500 に接続するための通信インタフェースである。SAN インタフェース 219 としては、ファイバチャネルプロトコルに対応したネットワークカードが用いられる。通信ポート 2191 は、情報処理装置 200 の SAN 500 への接続口であり、例えば、ファイバチャネルの通信規格における「N\_\_Port」や「NL\_\_Port」である。通信ポート 2191 には、SAN 500 上のノードを特定するためのネットワークアドレスである WWN が付与される。

#### 【0017】

図 1 に示すストレージ制御装置 100 は、チャンネル制御部 1 乃至 2（110）、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140、接続部 150 等を備えて構成される。ストレージ制御装置 100 は、情報処理装置 1 乃至 2（200）との間の通信に関する機能を提供するチャンネル制御部 1 乃至 2（110）を備えている。チャンネル制御部 1（110）は、情報処理装置 1 乃至 2（200）との間で LAN 400 を介した通信に関する機能を提供する。チャンネル制御部 2（110）は、情報処理装置 1 乃至 2（200）との間で SAN 500 を介した通信に関する機能を提供する。チャンネル制御部 1（110）とチャンネル制御部 2（110）とは、互いにバスライン 105 によって通信可能に接続されている。チャンネル制御部 1（110）とチャンネル制御部 2（110）とは、バスライン 105 を介して互いにデータ信号や割込信号等を送信もしくは受信することができる。以下ではチャンネル制御部 1（110）を CHN とも称する。また、チャンネル制御部 2（110）を CHF とも称する。

#### 【0018】

図 3 にチャンネル制御部 1（CHN）110 のハードウェア構成を示している。

チャンネル制御部 1 (CHN) 110 は、ネットワークインタフェース部 111、CPU 112、メモリ 113、入出力制御部 114、通信ポート 117、等を備えている。ネットワークインタフェース部 111 は、LAN 400 に接続するための通信インタフェースである。通信ポート 117 には、LAN 400 に接続するための通信線が接続される。

#### 【0019】

チャンネル制御部 1 (CHN) 110 では、オペレーティングシステムが動作している。また、オペレーティングシステム上では、各種のアプリケーションソフトウェアが動作している。オペレーティングシステムやアプリケーションソフトウェアの機能は、CPU 112 がメモリ 113 にロードされるプログラムが実行されることにより実現される。前記プログラムは、例えば、ディスクドライブ 300 や NVRAM 115 等に記憶されている。

#### 【0020】

オペレーティングシステムは、例えば、UNIX (登録商標) 系や Windows (登録商標) 系のオペレーティングシステムでありファイルシステムを備えている。また、オペレーティングシステム上では、NFS (Network File System)、CIFS (Common Internet File System) 等のファイルシステムプロトコルが動作しており、情報処理装置 1 乃至 2 からファイル指定によるデータ入出力要求 (以下、ファイルアクセス要求と称する) を受け付ける。すなわち、チャンネル制御部 1 (CHN) 110 は、情報処理装置 1 乃至 2 等の LAN 400 に接続する装置に対して NAS (Network Attached Storage) として機能している。

#### 【0021】

入出力制御部 114 は、ディスク制御部 140 やキャッシュメモリ 130、共有メモリ 120、との間でデータやコマンドの授受を行う。入出力制御部 114 は、I/O プロセッサ (Input/Output) 119 及び NVRAM (Non Volatile RAM) 115 を備えている。I/O プロセッサ 119 は、例えば、1 チップのマイコンで構成される。I/O プロセッサ 119 は、例えば、DMA (Direct Memory Access) プロセッサである。I/O プロセッサ 119 は、上記データやコマンドの授受を制御し、CPU 112 と接続部 (スイッチ) との間の通信を中継して

いる。NVRAM115は、I/Oプロセッサ119の制御を司るプログラムが格納される不揮発性メモリである。

#### 【0022】

図4にチャネル制御部2 (CHF) 110のハードウェア構成を示している。チャネル制御部2 (CHF) 110は、ネットワークインタフェース部111、メモリ113、入出力制御部114、I/Oプロセッサ119、NVRAM115、通信ポート117を備える。

#### 【0023】

チャネル制御部2 (CHF) 110は、CPU112を備えていない点でチャネル制御部1 (CHN) 110と構成が異なる。チャネル制御部2 (CHF) 110ではファイルシステムは動作しておらず、情報処理装置1乃至2 (200) からファイル指定ではなくブロック指定によるデータ入出力要求（以下、ブロックアクセス要求と称する）を受け付ける。また、チャネル制御部1 (CHN) 110がLAN400を介して行われる通信に関する機能を提供するのに対し、チャネル制御部2 (CHF) 110はSAN500を介して行われる通信に関する機能を提供する。

#### 【0024】

ネットワークインタフェース部111は、SAN500に接続するための通信インタフェースである。ネットワークインタフェース部111は、2つの通信ポート117を備えている。各通信ポート117には、SAN500に接続するための通信線が接続される。各通信ポート117には、それぞれSAN500におけるネットワークアドレスが付与される。例えば、SAN500の通信プロトコルがファイバチャネルである場合には、前記ネットワークアドレスは、WWN (World Wide Name) であり、各通信ポート117にはそれぞれ異なるWWNが付与される。

#### 【0025】

入出力制御部114は、ディスク制御部140、キャッシュメモリ130、共有メモリ120、とそれぞれとの間でデータやコマンドの授受を行う。入出力制御部114は、I/Oプロセッサ (Input/Output) 119及びNVRAM (Non

● Volatile RAM) 115を備えている。I/Oプロセッサ119は、例えば、1チップのマイコンで構成される。I/Oプロセッサ119は、例えば、DMA (Direct Memory Access) プロセッサである。I/Oプロセッサ119は、上記データやコマンドの授受を制御し、CPU112と接続部(スイッチ)150との間の通信を中継する。NVRAM115は、I/Oプロセッサ119の制御を司るプログラムが格納される不揮発性メモリである。

#### 【0026】

ディスク制御部140は、ディスクドライブ300の制御を行う。ディスク制御部140は、例えば、チャンネル制御部110が情報処理装置200から受信したデータ書き込みコマンドに従ってディスクドライブ300へデータの書き込みを行う。図5にディスク制御部140のハードウェア構成を示している。ディスク制御部140は、インタフェース部141、メモリ143、CPU142、NVRAM144、等を備えている。インタフェース部141は、接続部150を介してチャンネル制御部1乃至2(110)との間での通信や、共有メモリ120、キャッシュメモリ130へのアクセスを行うための通信インタフェースや、ディスクドライブ300との間で通信を行うための通信インタフェースを備えている。

#### 【0027】

CPU142は、ディスク制御部140全体の制御を司ると共に、チャンネル制御部1乃至2(110)やディスクドライブ300との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、例えば、ディスクドライブ300の制御やRAIDの機能に関する各種制御等がある。NVRAM144は、CPU142の制御を司るプログラムを格納する不揮発性メモリである。

#### 【0028】

図1に示すように、チャンネル制御部1乃至2(110)、共有メモリ120、キャッシュメモリ130、ディスク制御部140は、それぞれ接続部150を介して互いに通信可能に接続されている。接続部150は、例えば、高速スイッチ



ングによりデータ伝送を行う超高速クロスバススイッチなどのスイッチである。接続部150は、スイッチングに関する制御を行うための制御プロセッサ（不図示）を備えている。共有メモリ120及びキャッシュメモリ130は、チャンネル制御部1乃至2（110）、ディスク制御部140により共有されるメモリである。共有メモリ120は、主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ130は、主にデータを記憶するために利用される。

### 【0029】

===基本動作===

次に、情報処理装置1乃至2（200）から送信されてくるデータ書き込み要求やデータ読み出し要求などのデータ入出力要求を受信した場合におけるストレージ制御装置100の基本動作について説明する。

### 【0030】

まず、情報処理装置1乃至2（200）からストレージ制御装置100に対してデータ書き込み要求が送信された場合について説明する。ストレージ制御装置100のチャンネル制御部1乃至2（110）は、情報処理装置1乃至2（200）から送られてくるデータ書き込み要求を受信すると、データ書き込みコマンドを共有メモリ120に書き込むと共に、情報処理装置1乃至2（200）から受信した書き込みデータをキャッシュメモリ130に書き込む。チャンネル制御部1乃至2（110）は、キャッシュメモリ130に対するデータの書き込みが完了すると、情報処理装置200に書き込み完了報告を送信する。このように情報処理装置200への完了報告は、ディスクドライブ300への実際のデータの書き込み動作とは非同期に行われる。ディスク制御部140はリアルタイム（例えば、一定の時間間隔で）に共有メモリ120の内容を監視している。ディスク制御部140は、上記監視により共有メモリ120にデータ書き込みコマンドが書き込まれていることを検知すると、キャッシュメモリ130から書き込み対象となるデータ（以下、書き込みデータと称する）を読み出して、読み出した書き込みデータをディスクドライブ300に書き込む。以上のようにしてデータ書き込み要求に対応したディスクドライブ300へのデータの書き込みが行われる。

**【0031】**

次に、情報処理装置 200 からストレージ制御装置 100 に対してデータ書き込み要求が送信された場合におけるストレージ制御装置 100 の基本的な動作について説明する。ストレージ制御装置 100 は、情報処理装置 200 から送られてくるデータ読み出し要求を受信すると、この要求に対応するデータ読み出しコマンドをディスク制御部 140 に送出する。なお、チャンネル制御部 1 乃至 2 (110) からディスク制御部 140 へのデータ読み出しコマンドの伝達は、共有メモリ 120 を介して行われることもある。

**【0032】**

ディスク制御部 140 は、チャンネル制御部 1 乃至 2 (110) からデータ読み出しコマンドを受領すると、そのコマンドに指定されている読み出し対象のデータをディスクドライブ 300 から読み出して、読み出したデータをキャッシュメモリ 130 に書き込む。ディスク制御部 140 は、キャッシュメモリ 130 へのデータ転送が完了すると、その旨をチャンネル制御部 1 乃至 2 (110) に通知する。そして前記通知を受信したチャンネル制御部 1 乃至 2 (110) は、キャッシュメモリ 130 に記憶されている読み出し対象のデータを情報処理装置 1 乃至 2 (200) に転送する。

**【0033】**

上述したように、チャンネル制御部 1 (CHN) 110 は、情報処理装置 1 乃至 2 (200) からデータ入出力要求として LAN 400 を介してファイル指定によるファイルアクセス要求を受け付ける。一方、チャンネル制御部 2 (CHF) 110 は、情報処理装置 1 乃至 2 (200) からデータ入出力要求として SAN 500 を介してブロック指定によるブロックアクセス要求を受け付ける。

**【0034】**

=== 障害監視と障害対応 ===

チャンネル制御部 1 (CHN) 110 と情報処理装置 1 乃至 2 (200) とにおいては、これら装置間で互いに動作状態を監視し合う機能を提供するアプリケーションソフトウェアが動作している。なお、近年、複数のコンピュータを含んで構成されるシステムにおいては、可用性 (HA (High Availability)) の向上

、負荷分散（ロードバランシング）による処理効率の向上、同一処理の並列実行による信頼度を向上等を目的として、いわゆるクラスタの仕組みが導入されることが多い。このような場合には、お互いに他の装置の動作状態を監視する上記の機能は、クラスタの機能を実現するためのアプリケーションソフトウェア（以下、クラスタソフトウェア 160 と称する）の機能として提供される。

#### 【0035】

クラスタソフトウェア 160 は、LAN 400 を介してお互いにハートビートメッセージを送受信することで、相手方装置が正常に動作しているかどうかを監視している。クラスタソフトウェア 160 は、期待される時刻に相手方の装置から送られてくるハートビートメッセージを受信できている場合には、当該相手方装置は正常に動作しているものと判断する。また、クラスタソフトウェア 160 は、期待される時刻に相手の装置から送られてくるハートビートメッセージを受信できていない場合、すなわち、ハートビートメッセージが途絶している場合には、相手装置に何らかの障害が発生しているものと判断する。なお、上記にいう期待される時刻は、ハートビートメッセージがストレージ制御装置 100、情報処理装置 200、LAN 400 等に与える負荷や障害検知に要求される迅速性等を考慮して、適切な値に設定される。チャンネル制御部 1（200）において動作しているクラスタソフトウェア 160 は、ハートビートメッセージを送信してくる相手装置側の IP アドレスを記憶している。クラスタソフトウェア 160 は、ハートビートメッセージが送信されてこない情報処理装置 200 が存在する場合には、その情報処理装置 200 の IP アドレスを特定する機能を備えている。なお、他の装置の動作状態を監視する仕組みは、ハートビートメッセージを用いる上述のものに限られない。

#### 【0036】

ストレージ制御装置 100 は、以上に説明したクラスタソフトウェア 160 により提供される障害監視の仕組みによって、ある情報処理装置 200 からのハートメッセージが途絶していることを検知した場合には、その情報処理装置 200（以下、障害中の情報処理装置 200 と称する）からチャンネル制御部 2（110）に対して行われるデータ入出力要求に応じて実行される処理を制限するための

処理を実行する。ここでデータ入出力要求に応じて実行される処理を制限するための処理とは、例えば、通常の運用状態においてそのデータ入出力要求に対応して実行されるようにされている処理を実行しないようにすることをいう。具体的には、例えば、チャンネル制御部 2 (110) が障害中の情報処理装置 200 から送信されたデータ入出力要求を受信した場合に、そのデータ入出力要求に対応する処理を実行しないようにする処理、前記データ入出力要求に応じて出力される前記制御信号を出力しないように制御する処理、またはそのデータ入出力要求に対する情報処理装置 200 への受信通知を送信しないようにする処理をいう。以下では、チャンネル制御部 1 (CHN) 110 が、ある情報処理装置 200 からのハートビートメッセージが途絶していることを検知した場合に、前記情報処理装置 200 からチャンネル制御部 2 (CHF) 110 に対して送信されたデータ入出力要求に応じて実行される処理を制限する処理について説明する。

#### 【0037】

図 6 は、ある情報処理装置 200 からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置 200 からチャンネル制御部 2 (CHF) 110 に対して行われるデータ入出力要求に応じて実行される処理が制限される処理の一例を説明するフローチャートである。チャンネル制御部 1 (CHN) 110 で動作するクラスタソフトウェア 160 は、ある情報処理装置 200 において障害が発生し (S610)、その情報処理装置 200 からのハートビートメッセージが途絶していることを検知すると (S611)、障害中の情報処理装置 200 に付与されている SAN 500 上のネットワークアドレス (WWN) を検索する。ここでこの検索は、障害中の情報処理装置 200 に付与されている IP アドレスを特定し、その IP アドレスを検索キーとして、図 7 に示す情報処理装置-WWN 対応管理テーブル 700 を参照することにより行われる。情報処理装置-WWN 対応管理テーブル 700 は、チャンネル制御部 1 (CHN) 110 のメモリ 113、NVRAM 115、もしくは、ディスクドライブ 300 に記憶されているテーブルである。情報処理装置-WWN 対応管理テーブル 700 には、各情報処理装置 200 に付与されている LAN 400 上の IP アドレスと、情報処理装置 200 に付与されている SAN 500 上の WWN との対応づけが登録されている。

クラスタソフトウェア 160 は、検索した WWN をメモリ 113 に書き込む (S612)。

#### 【0038】

次にクラスタソフトウェア 160 は、メモリ 113 上に管理されている書き込み指示フラグを ON に設定する (S613)。チャンネル制御部 1 (CHN) 110 の I/O プロセッサ 119 は、書き込み指示フラグの内容をリアルタイムに監視している。チャンネル制御部 1 (CHN) 110 の I/O プロセッサ 119 は、書き込み指示フラグが ON になっていることを検知すると (S614: YES)、メモリ 113 に書き込まれている WWN を共有メモリ 120 に転送する (S615)。ここで共有メモリ 120 に転送された前記 WWN は、共有メモリ 120 に確保されている記憶領域である WWN 書き込みエリア 1201 に書き込まれる。

#### 【0039】

一方、チャンネル制御部 2 (CHF) 110 の I/O プロセッサ 119 は、共有メモリ 120 の WWN 書き込みエリア 1201 の内容をリアルタイムに監視している (S616)。I/O プロセッサ 119 は、前記監視により新たな WWN が WWN 書き込みエリア 1201 に書き込まれていることを検知すると (S617: YES)、その WWN を読み出して、チャンネル制御部 2 (CHF) 110 のメモリ 113 に転送する (S618)。これによりチャンネル制御部 2 (CHF) 110 のメモリ 113 には、障害中の情報処理装置 200 の WWN がリアルタイムに管理される。

#### 【0040】

チャンネル制御部 2 (CHF) 110 の I/O プロセッサ 119 は、チャンネル制御部 2 (CHF) 110 のメモリ 113 に前記 WWN が記憶されている場合には、情報処理装置 1 乃至 2 (200) から SAN 500 を介して送信されてくるデータ入出力要求についてそのデータ入出力要求に応じて実行される処理を制限するかどうかの判定を行う。そして、チャンネル制御部 2 (CHF) 110 は、メモリ 113 に記憶されている WWN と同じ WWN が付与されている情報処理装置 200 からデータ入出力要求が送信されてくると、そのデータ入出力要求についての処理を制限する (S619)。

#### 【0041】

ここでデータ入出力要求がメモリ 113 に記憶されている WWN が付与された情報処理装置 200 から送信されたものであるかどうかの判断は、メモリ 113 に記憶されている WWN と、送信元の通信ポート 2191 を特定するためにデータ入出力要求に付帯して送られてくる WWN とを対照することにより行われる。

#### 【0042】

以上の仕組みによれば、情報処理装置 200 において障害が発生した場合には、その情報処理装置 200 から送信されたデータ入出力要求に関する処理が制限される。これにより、情報処理装置 200 から送られてくる異常なコマンドや異常なデータによりディスクドライブ 300 に記憶されるデータが破損したり、データに管理上の不整合が生じたりすることを防ぐことができ、ストレージシステムの信頼性や可用性を向上させることができる。

#### 【0043】

なお、情報処理装置 200 の障害が復旧した場合には、クラスタソフトウェア 160 の機能により、もしくは、オペレータ等による手動操作により、該当の情報処理装置 200 の WWN が、WWN 書き込みエリア 1201 から削除される。そして、I/O プロセッサ 119 は、前記監視において、チャンネル制御部 2 (110) のメモリ 113 に記憶されている WWN が、共有メモリ 120 の WWN 書き込みエリア 1201 から削除されている場合には、チャンネル制御部 2 (110) のメモリ 113 からその WWN を削除する。これにより情報処理装置 200 の障害が復旧して正常に動作を開始した場合には、制限されていた該当の情報処理装置 200 から送信されたデータ入出力要求についての処理が自動的に再開される。

#### 【0044】

なお、例えば、WWN 書き込みエリア 1201 に記憶されている各 WWN に対応させたフラグ（以下、抑止解除フラグと称する）を共有メモリ 120 に管理し、チャンネル制御部 2 (CHF) 110 が抑止解除フラグの内容に応じてデータ入出力要求に応じた処理を制限する処理を実行するかどうかを決定するようにしてもよい。このように抑止解除フラグを用いる場合には、例えば、各チャンネル制御部 2 (CHF) が参照したかどうかを示す情報を共有メモリ 120 に管理してお

き、WWNを参照すべき全てのチャンネル制御部2（CHF）110が前記WWNを参照したことを確認した後に前記WWNをWWN書き込みエリア1201から削除するようにしてもよい。このようにすることで、チャンネル制御部2（CHF）110が複数存在する場合において各チャンネル制御部2（CHF）110が個別に上記データ入出力要求に応じた処理を制限する必要があるかどうかを判断できるようにすることができる。

#### 【0045】

図8はある情報処理装置200からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置200からチャンネル制御部2（110）に対して送信されてくるデータ入出力要求に応じて実行される処理が制限される処理の他の実施の形態を説明するフローチャートである。チャンネル制御部1（CHN）110で動作するクラスタソフトウェア160は、ある情報処理装置200において障害が発生し（S810）、その情報処理装置200からのハートビートメッセージが途絶していることを検知すると（S811）、いずれかの情報処理装置200に障害が発生していることを示す割込信号をバスライン105を介してチャンネル制御部2（110）に出力する（S812）。また、クラスタソフトウェア160は、障害中の情報処理装置200に付与されているSAN500上のネットワークアドレス（WWN）を検索する。ここでこの検索は、障害中の情報処理装置200に付与されているIPアドレスを特定し、そのIPアドレスを検索キーとして、情報処理装置－WWN対応管理テーブル700を参照することにより行われる。クラスタソフトウェア160は、検索したWWNをメモリ113に書き込む（S813）。

#### 【0046】

次にクラスタソフトウェア160は、メモリ113上に管理されている書き込み指示フラグをONに設定する（S814）。チャンネル制御部1（CHN）110のI/Oプロセッサ119は、書き込み指示フラグの内容をリアルタイムに監視している。チャンネル制御部1（CHN）110のI/Oプロセッサ119は、書き込み指示フラグがONになっていることを検知すると（S815:YES）、メモリ113に書き込まれているWWNを共有メモリ120に転送する（S815）。ここで共

有メモリ 120 に転送された前記 WWN は、共有メモリ 120 に確保されている記憶領域である WWN 書き込みエリア 1201 に書き込まれる。

#### 【0047】

一方、チャネル制御部 2 (CHF) 110 の I/O プロセッサ 119 は、チャネル制御部 1 (CHN) 110 から出力された割込信号を受信したのに応じて (S817) 共有メモリ 120 にアクセスし (S818)、WWN 書き込みエリア 1201 に書き込まれている新たな WWN をチャネル制御部 2 (CHF) 110 のメモリ 113 に記憶する (S819)。これによりチャネル制御部 2 (CHF) 110 のメモリ 113 には、障害中の情報処理装置 200 の WWN がリアルタイムに管理される。

#### 【0048】

チャネル制御部 2 (CHF) 110 の I/O プロセッサ 119 は、チャネル制御部 2 (CHF) 110 のメモリ 113 に前記 WWN が記憶されている場合には、情報処理装置 1 乃至 2 (200) から SAN 500 を介して送信されてくるデータ入出力要求についてそのデータ入出力要求に応じて実行される処理を制限するかどうかの判定を行う。そして、チャネル制御部 2 (CHF) 110 は、メモリ 113 に記憶されている WWN と同じ WWN が付与されている情報処理装置 200 からデータ入出力要求が送信されてきた場合にはそのデータ入出力要求に応じて実行される処理を制限する (S820)。ここでデータ入出力要求がメモリ 113 に記憶されている WWN が付与された情報処理装置 200 から送信されたものであるかどうかの判断は、メモリ 113 に記憶されている WWN と、送信元の通信ポート 2191 を特定するためにデータ入出力要求に付帯して送られてくる WWN とを対照することにより行われる。

#### 【0049】

以上の仕組みによれば、チャネル制御部 1 (CHN) 110 が、情報処理装置 200 における障害を検知した場合に、前記情報処理装置 200 から送信されたデータ入出力要求が着信されるチャネル制御部 2 (CHF) 110 により前記データ入出力要求に応じて実行される処理が制限されるように制御される。これにより障害が生じている情報処理装置 200 から送られてくる異常なデータ入出力



要求に応じてチャネル制御部 2 (CHF) 110 が異常な処理を行うことによりディスクドライブ 300 に記憶される（もしくは、記憶されている）データが破損したり、データに管理上の不整合が生じたりすることが防止され、ストレージシステムの信頼性や可用性を向上させることができる。また、チャネル制御部 1 (110) からチャネル制御部 2 (110) に対して情報処理装置 200 で障害が発生した場合にバスライン 105 を介して割込信号が出力され、チャネル制御部 2 (CHF) 110 は前記割込信号が入力されたのに応じて共有メモリ 120 にアクセスするので、リアルタイムに共有メモリ 120 にアクセスする上述の図 6 に示す仕組みに比べてチャネル制御部 2 (CHF) 110 の負荷も少なくて済む。

#### 【0050】

また、このようにストレージ制御装置 100 がその内部に複数のチャネル制御部 110 を備え、そのうちの少なくとも一つのチャネル制御部 110 が、情報処理装置 200 における障害を検知する仕組みを提供することができる構成である場合には、ストレージ制御装置 100 内部の通信により障害中の情報処理装置 200 から送信されたデータ入出力要求に応じて実行される処理を制限する仕組みを、ストレージ制御装置 100 の内部に持たせることが可能であり、信頼性及び可用性に優れたストレージ制御装置を提供することが可能となる。

#### 【0051】

=== 高速アクセス制御 ===

次に本発明の他の実施の形態に係るファイルの高速アクセス制御について説明する。本実施の形態に係るファイルの高速アクセス制御は、情報処理装置 200 から、ディスクドライブ 300 に記憶されているファイルデータに対して、SAN 500 を介したブロック単位の高速なデータアクセスを行うための制御である。図 1 に示すように、情報処理装置 1 (200) は、LAN 400 を介してチャネル制御部 1 (CHN) 110 と接続されている。また情報処理装置 1 (200) は、SAN 500 を介してチャネル制御部 2 (CHF) 110 とも接続されている。これにより情報処理装置 1 (200) は、チャネル制御部 1 (CHN) 110 を通じてディスクドライブ 300 に記憶されているファイルデータをアク

セスすることもできるし、チャンネル制御部 2 (CHF) 110 を通じて上記同一データをアクセスすることもできる。但し、チャンネル制御部 1 (CHN) 110 を介してアクセスする場合はファイル単位でのアクセスとなるのに対し、チャンネル制御部 2 (CHF) 110 を介してアクセスする場合はブロック単位でのアクセスとなる。

#### 【0052】

通常、情報処理装置 1 (200) が、チャンネル制御部 1 (CHN) 110 を介してディスクドライブ 300 に記憶されているデータにアクセスする場合には、チャンネル制御部 1 (CHN) 110 に対してファイル名を指定したファイルアクセス要求を行うが、本実施の形態に係るファイルの高速アクセス制御によりディスクドライブ 300 に記憶されているデータにアクセスする場合には、情報処理装置 1 (200) は、まずチャンネル制御部 1 (CHN) 110 に対してファイル名を指定して、ファイルについての記憶装置の記憶領域上の記憶位置を特定する情報であるメタデータ (ファイル管理情報) の要求 (リクエスト) を行う。メタデータは、例えば、UNIX (登録商標) における「i-node」である。メタデータの要求を受け付けたチャンネル制御部 1 (CHN) 110 は、メモリ 113 またはキャッシュメモリ 130 に記憶されている当該ファイル名に対応するメタデータを読み出す。そして、読み出したメタデータを LAN 400 を介して情報処理装置 1 (200) に送信する。なお、メタデータはディスクドライブ 300 にも記憶されているので、チャンネル制御部 1 (CHN) 110 はディスクドライブ 300 からメタデータを読み出すようにすることもできる。

#### 【0053】

情報処理装置 (200) はメタデータを取得することにより、当該ファイルの記憶位置やデータサイズを知ることができる。情報処理装置 (200) はこれらの情報に基づいてファイルデータに対するブロックアクセス要求を生成する。そして当該ブロックアクセス要求を SAN 500 を介してチャンネル制御部 2 (CHF) 110 に対して送信する。

#### 【0054】

チャンネル制御部 2 (CHF) 110 は、ネットワークインタフェース部 111

により上記ブロックアクセス要求を受け付ける。そして I/O プロセッサ 119 は当該データの記憶位置とデータ長等を抽出し、上記ブロックアクセス要求に対応する I/O 要求を生成してディスク制御部 140 に出力する。このようにしてデータの読み出しや書き込み等が行われる。

#### 【0055】

SAN 500 は LAN 400 と比較して高速なデータ転送が可能なネットワークであるので、ディスクドライブ 300 に記憶されているファイルデータに高速にアクセスすることができる。

情報処理装置 1 (200) は、ディスクドライブ 300 からファイルデータを読み出す場合には、チャンネル制御部 2 (CHF) 110 に対して当該ファイルデータのアドレスとサイズを指定してブロック単位でのデータ読み出し要求を送信する。チャンネル制御部 2 (CHF) 110 はディスクドライブ 300 から読み出したデータを SAN 500 を介して情報処理装置 1 (200) に送信する。情報処理装置 1 (200) は、データをチャンネル制御部 2 (CHF) 110 から取得したら読み出し処理を終了する。なお、チャンネル制御部 1 (CHN) 110 からメタデータを取得する際に当該ファイルに対してロックを掛けていた場合には、チャンネル制御部 1 (CHN) 110 に対してロック解除要求を送信する。

#### 【0056】

一方、ディスクドライブ 300 にファイルデータを書き込む場合には、情報処理装置 1 (200) は、当該書き込みデータと共に書き込みデータのアドレスとサイズを指定してブロック単位でのデータ書き込み要求をチャンネル制御部 2 (CHF) 110 に対して送信する。チャンネル制御部 2 (CHF) 110 は当該書き込みデータをディスクドライブ 300 に書き込み、書き込み完了メッセージを情報処理装置 1 (200) に送信する。情報処理装置 1 (200) はチャンネル制御部 2 (CHF) 110 から書き込み完了のメッセージを受信したら、チャンネル制御部 1 (CHN) 110 に対してメタデータの更新を要求する。

#### 【0057】

本実施の形態に係るファイルの高速アクセス制御は、データサイズの大きなファイルにアクセスする場合に効果が大きい。データサイズの大きなファイルへの

アクセスを高速な SAN 500 を介して行うことにより、ファイルデータに対する読み出し、書き込みの時間を短縮できる。これは本実施の形態に係るストレージシステム 600 においては、ストレージ制御装置 100 のスロット内にチャンネル制御部 1 (CHN) 110、チャンネル制御部 2 (CHF) 110 を混在させて装着することが可能であり、チャンネル制御部 1 (CHN) 110 を介したデータアクセスとチャンネル制御部 2 (CHF) 110 を介したデータアクセスのそれぞれの特長をうまく利用することができる場合に実現できるのである。

#### 【0058】

本実施の形態にかかるファイルの高速アクセス制御は、情報処理装置 1 (200) からチャンネル制御部 1 (CHN) 110 に対するアクセスと、情報処理装置 1 (200) からチャンネル制御部 2 (CHF) 110 に対するアクセスとが、関連をもって行われる。上述した障害監視と障害対応の仕組みは、このような形態においても実現される。つまり、情報処理装置 1 (200) とクラスタを構成しているチャンネル制御部 1 (CHN) 110 が、情報処理装置 1 (200) において生じている障害を検知すると、チャンネル制御部 2 (CHF) 110 により前記データ入出力要求に応じて実行される処理が制限されるように制御される。すなわち、このように情報処理装置 1 (200) からチャンネル制御部 1 (CHN) 110 に対するアクセスと、情報処理装置 1 (200) からチャンネル制御部 2 (CHF) 110 に対するアクセスとが、関連をもって行われる場合に、上述した障害監視と障害対応の仕組みは、特に有用である。これにより、障害が生じている情報処理装置 1 (200) から送られてくる異常なデータ入出力要求に応じてチャンネル制御部 2 (CHF) 110 が異常な処理を行うことによりディスクドライブ 300 に記憶される（もしくは、記憶されている）データが破損したり、データに管理上の不整合が生じたりすることが防止され、高速アクセス制御が行われた場合におけるストレージシステムの信頼性や可用性を向上させることができる。

#### 【0059】

===ストレージシステムの他の形態===

図 9 は、ストレージシステムの他の形態を示している。このストレージシステ

ムは、ストレージ制御装置 1 0 0、ストレージ制御装置 1 0 0 に通信可能に接続しストレージ制御装置 1 0 0 に対してデータ入出力要求を送信する情報処理装置 1 乃至 5 ( 2 0 0 )、ストレージ制御装置 1 0 0 により制御されデータの記憶領域を提供する記憶デバイスである一台以上のディスクドライブ 3 0 0 等を含んで構成される。ストレージ制御装置 1 0 0 は、情報処理装置 1 乃至 5 ( 2 0 0 ) から送信されたデータ入出力要求を受信して、このデータ入出力要求に応じてディスクドライブ 3 0 0 に記憶されているデータの書き込みや読み出しに関する処理を行う。

#### 【 0 0 6 0 】

ストレージ制御装置 1 0 0 と情報処理装置 1 乃至 5 ( 2 0 0 ) とは、LAN ( Local Area Network ) 4 0 0 を介して通信可能に接続されている。LAN 4 0 0 を介して行われる通信は、例えば、TCP / IP プロトコルに従って行われる。また、情報処理装置 3 乃至 4 ( 2 0 0 ) とストレージ制御装置 1 0 0 とは、SAN ( Storage Area Network ) 5 0 0 を介して通信可能に接続されている。SAN 5 0 0 を介して行われる通信はファイバチャネルに従って行われる。情報処理装置 5 ( 2 0 0 ) とストレージ制御装置 1 0 0 とは、iSCSI、FICON (登録商標) や ESCON (登録商標)、ACONARC (登録商標)、FIBAR C (登録商標) 等の通信プロトコルに従って通信が行われる通信経路 5 5 0 を介して通信可能に接続されている。

#### 【 0 0 6 1 】

情報処理装置 1 乃至 5 ( 2 0 0 ) は、それぞれ CPU ( Central Processing Unit ) やメモリを備えるコンピュータである。情報処理装置 1 乃至 5 ( 2 0 0 ) は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等である。情報処理装置 1 乃至 5 ( 2 0 0 ) が備える CPU により各種プログラムが実行されることにより、情報処理装置 1 乃至 5 ( 2 0 0 ) が提供する様々な機能が実現される。また、情報処理装置 1 乃至 5 ( 2 0 0 ) では、オペレーティングシステムが動作している。また、オペレーティングシステム上では、各種のアプリケーションプログラムが動作しており、情報処理装置 1 乃至 5 ( 2 0 0 ) のうちの少なくともいずれかにおいては上述のクラスタソフトウェア

160が動作している。

#### 【0062】

ストレージ制御装置100は、情報処理装置1乃至5(200)との間の通信に関する機能を提供する、チャンネル制御部1乃至8(110)を備えている。このうちチャンネル制御部1乃至4(CHN)110は、情報処理装置1乃至3(200)との間でLAN400を介した通信に関する機能を提供する。なお、チャンネル制御部1乃至8(110)は、バスライン105によって互いに通信可能に接続されている。チャンネル制御部1乃至4(110)は、このバスライン105を介して互いにデータ信号や割込信号等を送信もしくは受信することができる。

#### 【0063】

チャンネル制御部1乃至4(110)は、第一の実施例で説明したチャンネル制御部1(CHN)110と同等のハードウェア及びソフトウェア構成を備える。また、チャンネル制御部5乃至8(CHF)110は、第一の実施例で説明したチャンネル制御部2(CHF)110と同等のハードウェア及びソフトウェア構成を備える。

#### 【0064】

チャンネル制御部1乃至4(CHF)110のうちの少なくともいずれかにおいては、上述のクラスタソフトウェア160が動作している。クラスタソフトウェア160は、ある情報処理装置200からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置200から他のチャンネル制御部1乃至8(110)に対して行われるデータ入出力要求に応じて実行される処理を制限させる、図6もしくは図8のフローチャートで説明した処理に相当する処理を実行する。なお、図9に示すストレージシステムの場合には、チャンネル制御部1(CHN)110がある情報処理装置200における障害を検知した場合にチャンネル制御部2(CHF)110に対してのデータ入出力要求に応じて実行される処理が制限されるだけでなく、当該チャンネル制御部1(CHN)110以外の他のチャンネル制御部1(CHN)110に対して送信されるデータ入出力要求に応じて実行される処理についても制限されるようにすることもできる。

#### 【0065】

ディスク制御部 1 乃至 4 (140) のハードウェア及びソフトウェア構成は、第一の実施例で説明したディスク制御部 140 と同等の構成である。チャンネル制御部 1 乃至 8 (110)、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 は、接続部 150 を介して通信可能に接続されている。接続部 150 は、例えば、高速スイッチングによりデータ伝送を行う超高速クロスバスイッチなどのスイッチである。共有メモリ 120 及びキャッシュメモリ 130 は、チャンネル制御部 110、ディスク制御部 140 により共有されるメモリである。共有メモリ 120 は、主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ 130 は、主にデータを記憶するために利用される。

#### 【0066】

ディスクドライブ 300 は、情報処理装置 200 に対して記憶領域を提供する。データは、ディスクドライブ 300 により提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリュームに記憶されている。ディスクドライブ 300 としては、例えば、ハードディスク装置や半導体記憶装置等、様々なものを用いることができる。情報処理装置 1 乃至 5 (200) からデータ入出力要求を受信した場合におけるストレージ制御装置 100 の基本的な動作は、第一実施例のストレージシステムにおける場合と基本的に同様である。

#### 【0067】

管理コンピュータ (SVP) 160 は、ストレージシステム 600 を保守・管理するためのコンピュータである。管理コンピュータ 160 は、ストレージ制御装置 100 の内部に設けられている LAN である内部 LAN 560 を介してチャンネル制御部 1 乃至 8 (110) 及びディスク制御部 1 乃至 4 (140) と接続している。

#### 【0068】

管理コンピュータ 160 を操作することにより、例えば、ディスクドライブ 300 の設定や、論理ボリュームの設定、チャンネル制御部 1 乃至 8 (110) において実行されるマイクロプログラムのインストール等を行うことができる。ここで、ディスクドライブ 300 の設定としては、例えば、ディスクドライブ 300 の増設や減設、RAID 構成の変更 (例えば RAID 1 から RAID 5 への変更

等)等を行うことができる。さらに管理コンピュータ160からは、ストレージシステム600の動作状態の確認や故障部位の特定、チャンネル制御部1乃至4 (CHN) 110で実行されるオペレーティングシステムやアプリケーションプログラムのインストール等の作業を行うこともできる。

#### 【0069】

管理コンピュータ160は、ストレージ制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。管理コンピュータ160は、ストレージ制御装置100及びディスクドライブ300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

#### 【0070】

チャンネル制御部1乃至4 (CHN) 110で動作しているオペレーティングシステム701上では、上述したクラスタソフトウェア160以外にも、RAID マネージャ、ボリュームマネージャ、ファイルシステムプログラム、NFS (Network File System) 等の様々なソフトウェアが動作している。このうち、RAID マネージャは、情報処理装置1乃至5 (200) のユーザやオペレータ等が、ディスク制御装置140に対してパラメータの設定や制御を行うためのソフトウェアである。設定されるパラメータの種類としては、例えば、RAID グループを構成するディスクドライブ300 (物理ディスク) を定義 (RAID グループの構成情報、ストライプサイズの指定など) するためのパラメータ、RAID レベル (例えば0, 1, 5) を設定するためのパラメータ等がある。ボリュームマネージャは、RAID 制御部740によって提供されるLUをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム703に提供する。1つの仮想化論理ボリュームは1以上の論理ボリュームによって構成される。ファイルシステムプログラムは、ネットワーク制御部702が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されている仮想化論理ボリューム上のアドレスとの対応づけを管理する。例えば、ファイルシステムプログラム703はファイルアクセス要求に指定されているファイル名に対応する仮想化論理ボリューム上のアドレスを特定する。また、NFS (Network File



System) 711は、NFS 711が動作するUNIX（登録商標）系の情報処理装置200からのファイルアクセス要求を受け付ける。

#### 【0071】

以上に説明した発明の実施の形態は、本発明の理解を容易にするためのものであり、本発明を限定するものではない。本発明は、その趣旨を逸脱することなく、変更、改良され得ると共に、本発明にはその等価物が含まれることは勿論である。

#### 【0072】

##### 【発明の効果】

本発明によれば、ストレージ制御装置の信頼性や可用性を向上させることができる。

##### 【図面の簡単な説明】

【図1】 本発明の一実施例によるストレージシステムの概略構成を示す図である。

【図2】 本発明の一実施例による情報処理装置のハードウェア構成を示す図である。

【図3】 本発明の一実施例によるチャンネル制御部1（CHN）のハードウェア構成を示す図である。

【図4】 本発明の一実施例によるチャンネル制御部2（CHF）のハードウェア構成を示す図である。

【図5】 本発明の一実施例によるディスク制御部のハードウェア構成を示す図である。

【図6】 本発明の一実施例による、データ入出力要求に応じて実行される処理が制限される際の処理を説明するフローチャートを示す図である。

【図7】 本発明の一実施例による情報処理装置－WWN対応管理テーブルを示す図である。

【図8】 本発明の一実施例による、データ入出力要求に応じて実行される処理が制限される際の処理の他の例を説明するフローチャートを示す図である。

【図9】 本発明の一実施例によるストレージシステムの他の形態を示す図

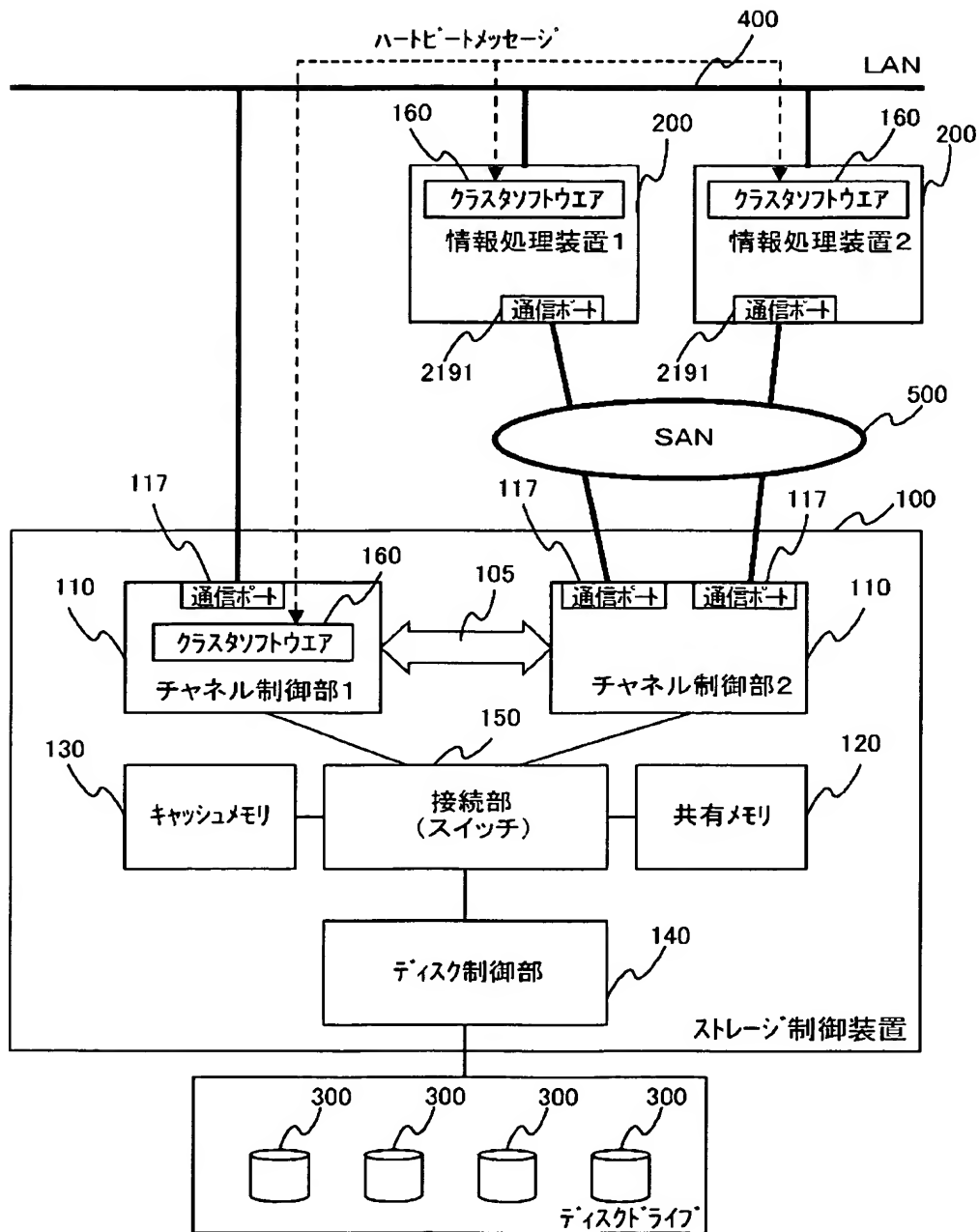
である。

【符号の説明】

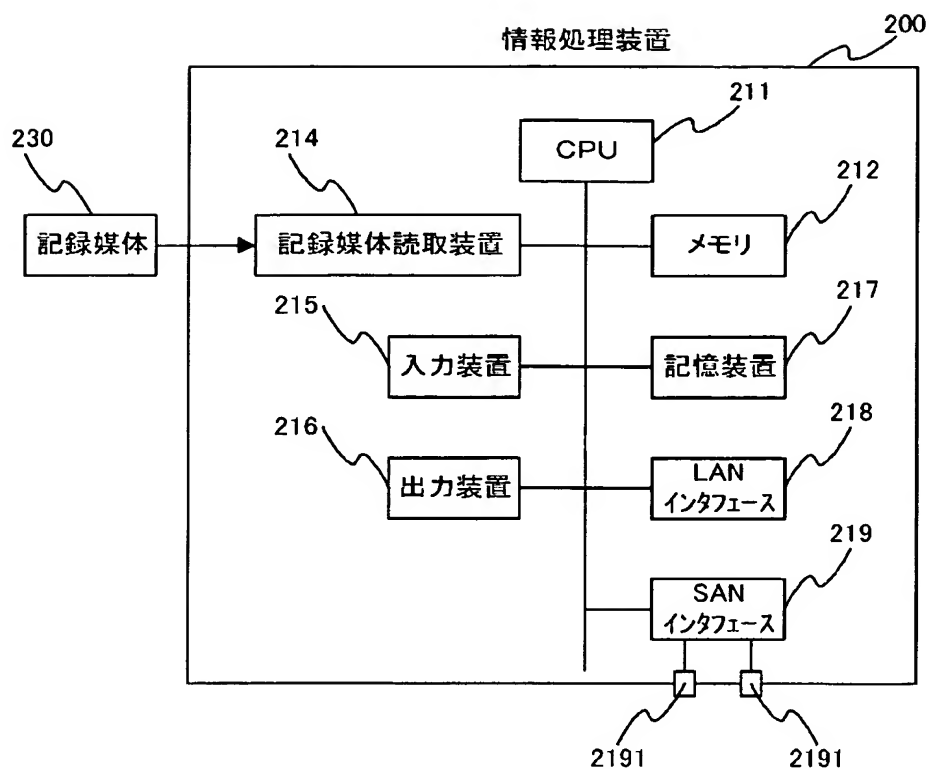
1 0 0    ストレージ制御装置  
1 0 5    バスライン  
1 1 0    チャネル制御部  
1 1 1    ネットワークインタフェース部  
1 1 2    C P U  
1 1 3    メモリ  
1 1 7    通信ポート  
1 1 9    I / O プロセッサ  
1 2 0    共有メモリ  
1 3 0    キャッシュメモリ  
1 5 0    接続部  
2 0 0    情報処理装置  
2 1 9 1    通信ポート  
3 0 0    ディスクドライブ  
4 0 0    L A N  
5 0 0    S A N

【書類名】 図面

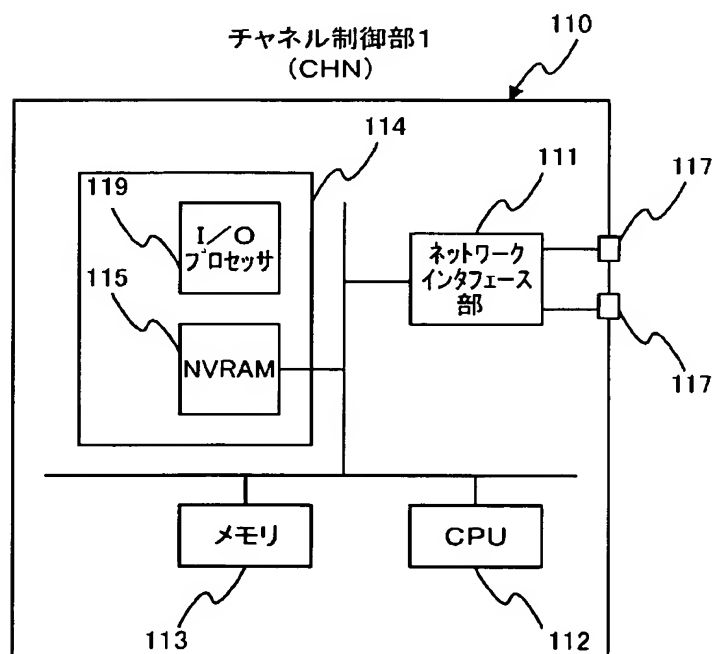
【図 1】



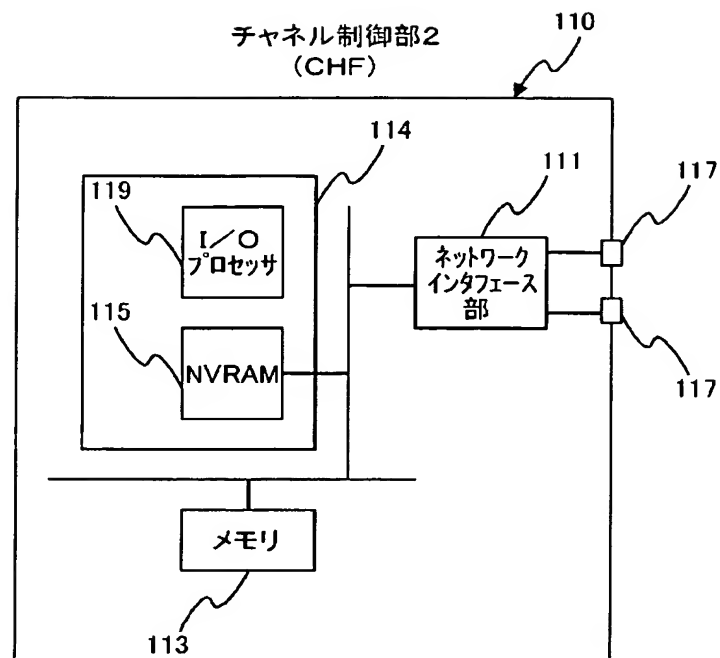
【図 2】



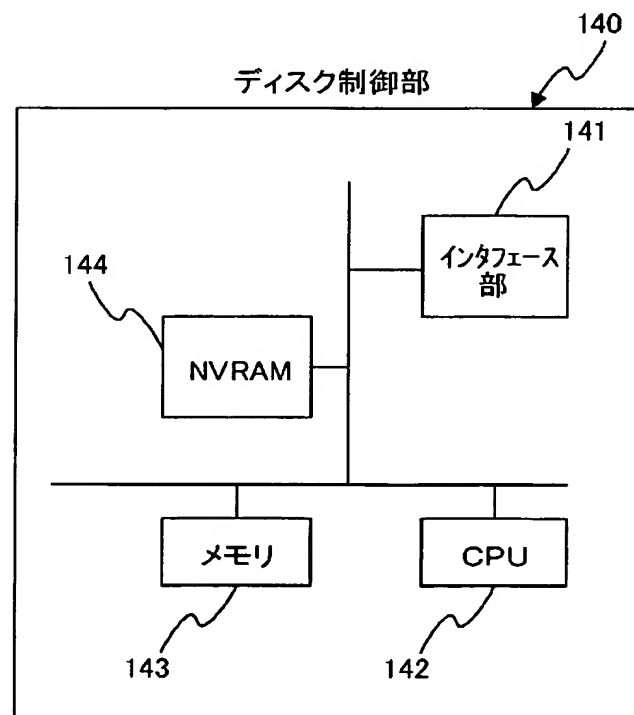
【図 3】



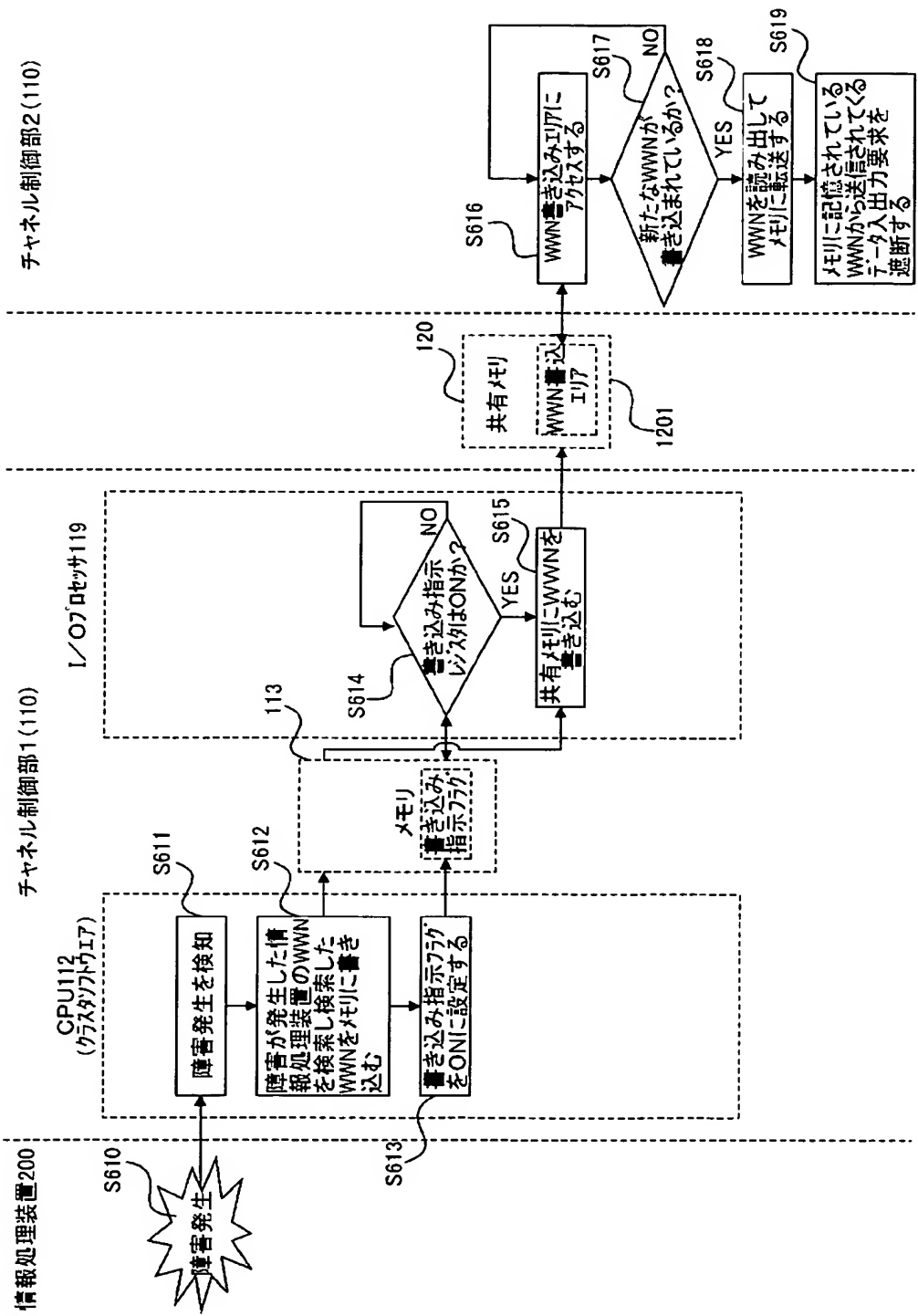
【図 4】



【図 5】



【図 6】



【図 7】

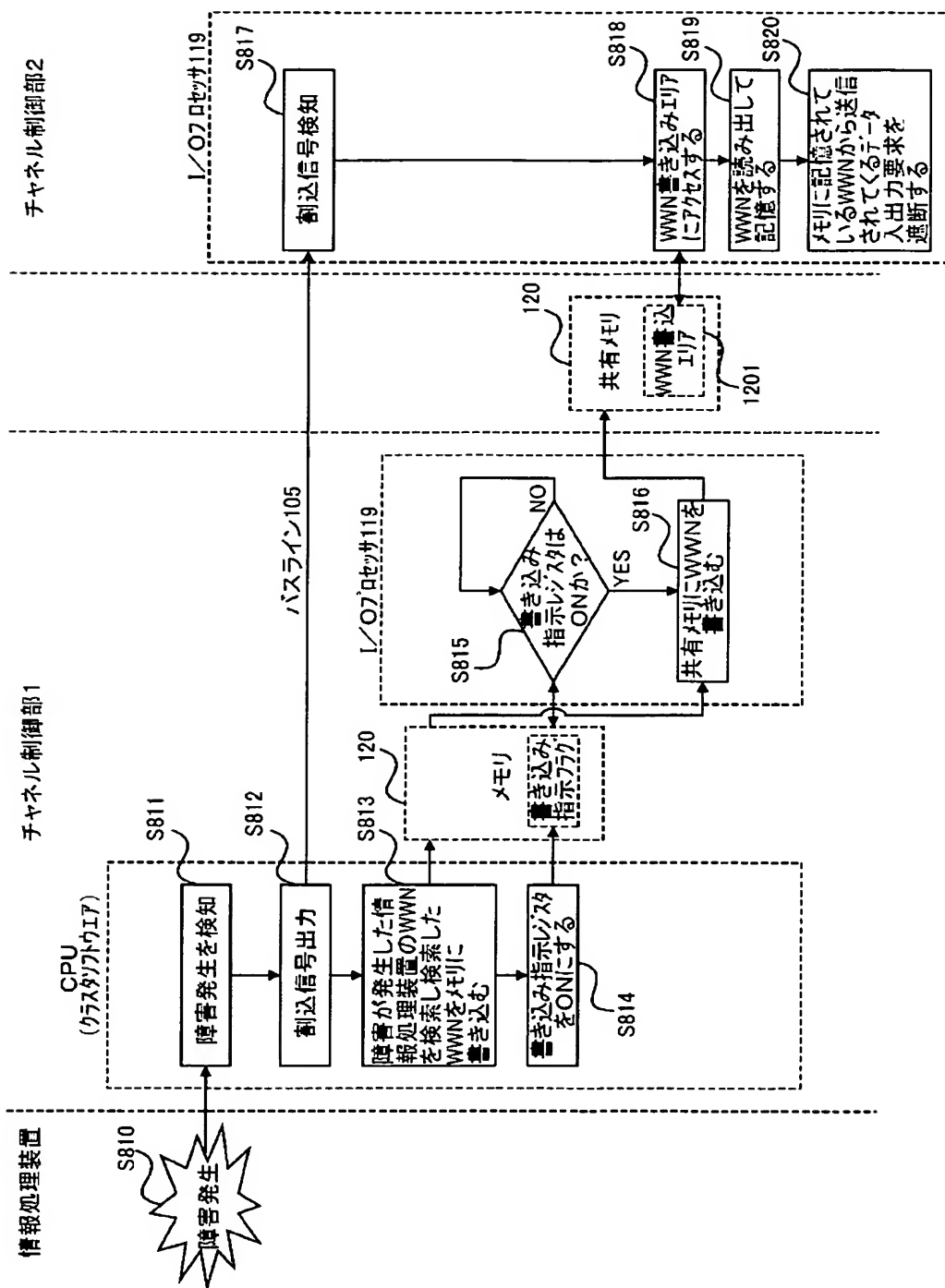
情報処理装置-WWN対応管理テーブル

700



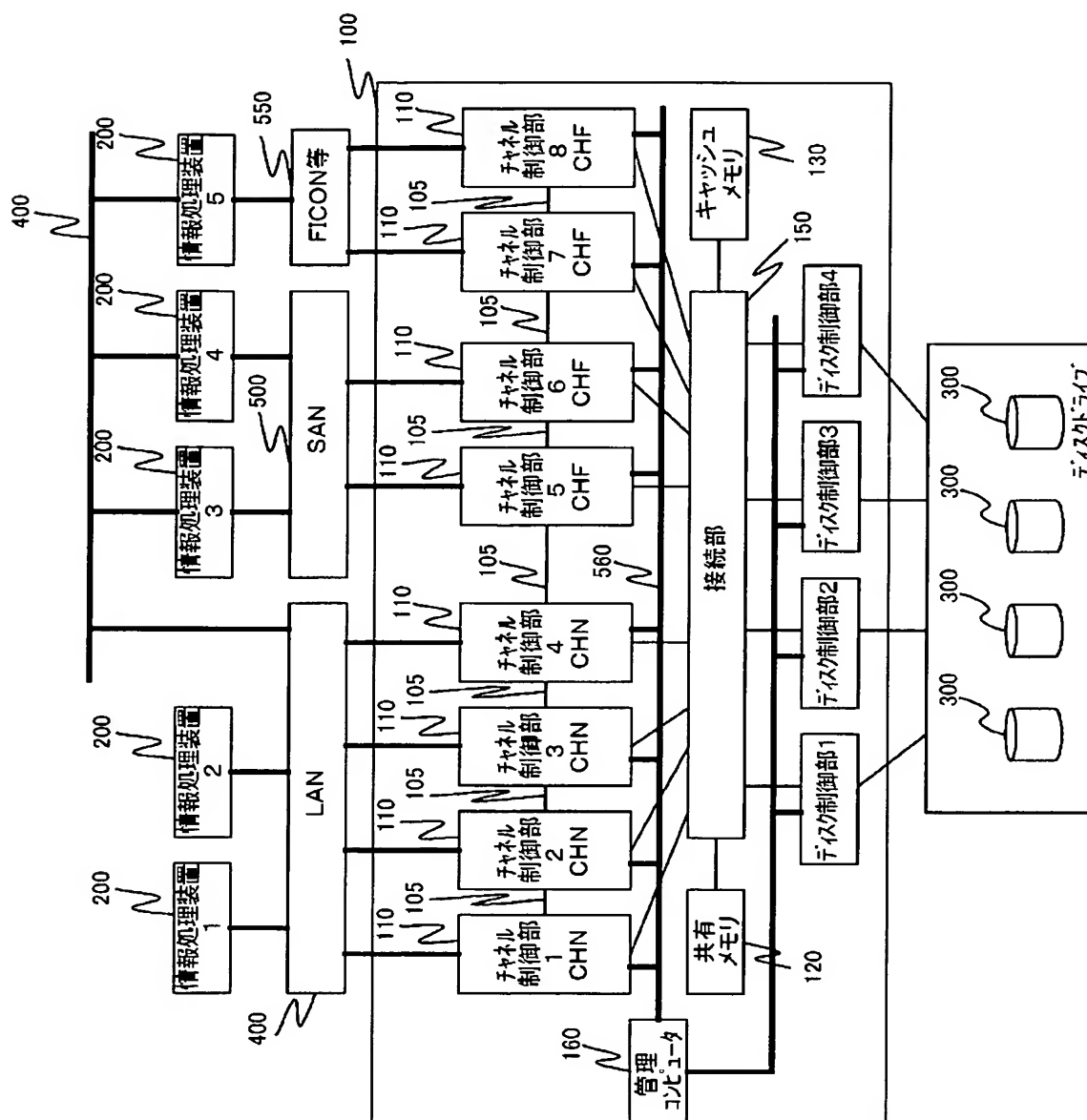
情報処理装置の IPアドレス	情報処理装置の通信ポート に付与されているWWN
192.168.0.1	10:00:00:00:00:00:01
192.168.0.2	10:00:00:00:00:00:02
192.168.0.3	12:00:00:00:00:00:01
192.168.0.4	12:00:00:00:00:00:02
.	.
.	.
.	.

【図 8】





【圖 9】



【書類名】 要約書

【要約】

【解決手段】 情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャンネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、第1の前記チャンネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、第1の前記チャンネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第2の前記チャンネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行するようにする。

【選択図】 図1

特願 2 0 0 3 - 1 5 8 2 7 1

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所